WHAT DRIVES RACIAL SEGREGATION? NEW EVIDENCE USING CENSUS MICRODATA*

Patrick Bayer*
Department of Economics
Yale University

Robert McMillan*
Department of Economics
University of Toronto

Kim Rueben*
Public Policy Institute of California

CES 02-26 October, 2002

Abstract

Residential segregation on the basis of race is widespread and has important welfare consequences. This paper sheds new light on the forces that drive observed segregation patterns. Making use of restricted micro-Census data from the San Francisco Bay Area and a new measurement framework, it assesses the extent to which the correlation of race with other household characteristics, such as income, education and immigration status, can explain a significant portion of observed racial segregation. In contrast to the findings of the previous literature, which has been hampered by serious data limitations, our analysis indicates that individual household characteristics *can* explain a considerable fraction of segregation by race. Taken together, we find that the correlation of race with other household attributes can explain almost 95 percent of segregation for Hispanic households, over 50 percent for Asian households, and approximately 30 percent for White and Black households. Our analysis also indicates that different factors drive the segregation of different races. Language explains a substantial proportion - more than 30 percent - of Asian and Hispanic segregation, education explains a further 20 percent of Hispanic segregation, while income is the most important non-race household characteristic for Black households, explaining around 10 percent of Black segregation.

* We would like to thank Fernando Ferreira (University of California – Berkeley) for outstanding research assistance as well as Jackie Chou and Pedro Cerdan for help assembling the data used in the paper. We would also like to thank Debbie Reed, Steve Ross, Jon Sonstelie, Chris Udry, Junfu Zhang and seminar participants at Johns Hopkins, UCLA, and Yale for helpful comments. We are grateful to the California Census Research Data Center for providing access to the data and to Ritch Milby in particular. All output included in this paper has been subject to thorough disclosure analysis by Census Bureau officials and meets stringent standards to safeguard confidentiality. Financial support from the Public Policy Institute of California is gratefully acknowledged. Please send correspondence to any of the authors - patrick.bayer@yale.edu, mcmillan@chass.utoronto.ca, or rueben@ppic.org.

Residential segregation on the basis of race and ethnicity is evident in cities throughout the United States. It is often accompanied by marked differences in the income, education and other socio-demographic characteristics of residents across racially segregated neighborhoods. These differences matter. As a growing body of evidence relating to neighborhood effects indicates, racial segregation worsens economic outcomes for individuals, especially those growing up in high-poverty neighborhoods isolated from the mainstream economy.[1]

The most natural explanation for racial segregation is that race itself is a fundamental driving force in the housing market, with segregation arising because households take account of the race of their neighbors when making their residential choices or because of centralized discrimination in the housing market. As observed by Schelling (1971) and many subsequent authors, it is entirely possible, however, that a number of alternative mechanisms only indirectly related to race have a sizeable impact on the level of racial segregation. Households might sort among neighborhoods based on their wealth or income; individuals might have tastes for characteristics of their neighbors that are correlated with race, such as speaking the same language; and information about desirable locations or jobs might flow through social networks that households are part of, leading like households to cluster in similar locations. In each of these examples, the correlation of race with other household characteristics such as income, wealth, education, language, or immigration status, provides an alternative explanation for racial segregation. Using unique and previously untapped micro-data drawn from the restricted-access version of the 1990 Census, the main goal of this paper is to provide new evidence on an old question – to what extent can the correlation of race with other sociodemographic characteristics explain the observed degree of residential segregation on the basis of race? To that end, we develop a measurement approach that allows us to take advantage of the underlying differences in the joint distribution of characteristics provided in the restricted data.

The existing empirical literature that attempts to understand the forces underlying segregation can be divided into two main categories. A number of studies have used data characterizing differences in the prices paid for comparable houses by households of different races to distinguish whether segregation arises because of centralized discriminatory practices or the decentralized residential location decisions made by the households of a metropolitan area, with preferences defined

[1] See Borjas (1995) and Cutler and Glaeser (1997) for recent evidence.

over the race of their neighbors.[2] As this central question indicates, the focus of these studies has been on factors directly linked to race. The second main line of research has attempted to explore whether sorting on the basis of other sociodemographic characteristics can explain the observed level of racial segregation. This line of research has been hampered by serious data limitations, forcing researchers using micro data to study sorting over large geographic areas such as counties (Gabriel and Rosenthal (1989)) or PUMAs, Census-defined areas made up of at least 100,000 people (Bajari and Kahn (2001)). In order to use data characterizing the racial composition of smaller geographic areas such as Census tracts or zip codes, researchers have generally had to use data that do not explicitly provide information on each household. Miller and Quigley (1990) and Harsman and Quigley (1995), for example, compare the degree of racial segregation in a metropolitan area to the degree of stratification on the basis of income and other household characteristics, concluding that sorting on the basis of these other characteristics can explain only a small amount of the racial segregation. Using tract-level data that classify individuals into income categories by race and education categories by race, Clark and Ware (1997) are able to explore how the segregation of households of each race varies over the income and education distributions, respectively. They find evidence that the equalization of education or income across race would decrease the segregation of each race, but again the nature of the data prevents the authors from exploring the role of other household characteristics such as wealth, language or immigration status or from examining the marginal impact of education or income.

In direct contrast to the data limitations previously faced by these authors, the recently available restricted-access Census data for 1990 match data for each household with its Census block - an area with approximately 100 residents. These data allow us to characterize the people of a major metropolitan area (over 240,000 households and 650,000 individuals in the San Francisco Bay Area) and their neighborhoods much more accurately than has been previously possible.[3] In particular we

[2] Notable papers in this line of research include King and Mieszkowski (1973), Schnare (1976), Yinger (1978), Schafer (1979), Follain and Malpezzie (1981), Chambers (1992), Kiel and Zabel (1996), and Cutler, Glaeser, and Vigdor (1999). These papers provide mixed evidence concerning whether Black households pay a premium for comparable housing, suggesting that the existence of such a premium may vary over time and location and by how well the researcher controls for unobserved neighborhood quality.

[3] The study of Borjas (1998), which is the work most closely related to our own, deserves special attention. This study uses a restricted version of the NLSY, generating neighborhood socio-demographics from the characteristics of other the individuals in the sample who reside in the same ZIP code for the 1979 wave of the survey. Because the NLSY is a national survey with a limited sample size, however, the socio-demographic composition of each ZIP code in the Borjas study is based on a very limited sample of the other individuals in the ZIP code. Relative to the Borjas study, these newly available restricted Census data

are able to examine the exposure of households of each race to every other race conditional on a wide set of household attributes, and so to characterize the marginal impact of a wide set of characteristics on the propensity of households of each race to live with other households of the same and other races. While the direct examination of the effect of household characteristics on the segregation patterns of households of each race is interesting in its own right, the primary goal of our analysis is to examine the extent to which the correlation of race with a series of observed sociodemographic characteristics can explain the observed degree of racial residential segregation. Making this distinction requires additional assumptions about the primitives of the sorting process. While it is important to emphasize that we are not modeling the underlying sorting process explicitly[4] and the counterfactual exercises that we carry out are not fully general equilibrium in nature, the counterfactual simulations that we conduct provide the best answer to date on the primary research question posed in this paper.

In line with all studies in the literature that have documented segregation patterns, our results indicate that households of each major racial/ethnic group in the Bay Area (Hispanic and Non-Hispanic Asian, Black and White) live in neighborhoods with an overrepresentation of households of the same racial or ethnic group, with Black households experiencing the most extreme level of segregation.[5] Again in line with previous findings, we find that the majority of racial segregation occurs below the level of county or PUMA and that a substantial amount of segregation occurs even within Census tracts. In overall terms, our analysis shows that household characteristics, including education, income, language, and immigration status, together conservatively explain almost 95 percent of segregation for Hispanic households, over 50 percent for Asian households, and approximately 30 percent for White and Black households. These numbers are much larger than those previously reported in the literature, emphasizing the value of the fine geographic detail of the micro data used in this study. Different factors drive the segregation of different races: language and

---

provide detailed information on the characteristics of a much wider sample of households observed at a lower level of aggregation. In turn, they give a richer view of the actual underlying socio-demographic composition of each neighborhood. It is also worth noting that the focus of Borjas (1998) is in studying the role of human capital externalities in segregation choices made by households of different ethnic groups rather than in exploring the role of a broad set of household characteristics.

[4] That task is carried out in related work - see Bayer, McMillan, and Rueben (2002).

[5] Massey and Denton (1987, 1989, 1993), Miller and Quigley (1990), Harsman and Quigley (1995), and Frey and Farley (1996) document segregation patterns, particularly between Black and White households, and how these have changed over time. These articles examine the level and changes in segregation patterns using various measures including dissimilarity indices and exposure indices. Massey and Denton (1998) and Reardon and Firebaugh (2002) examine a wide variety segregation measures and discuss the advantages and problems with each different measure.

immigration status explain a considerable fraction of segregation of Asian and Hispanic households, while income is the most important non-race household characteristic in explaining Black segregation. These results suggest that race itself is likely to be a far more important factor in driving the segregation of Black households than it is for Hispanic and, to a lesser extent Asian, households.

I DATA

Our analysis is conducted using an extensive new data set built around restricted Census microdata for 1990. These restricted Census data provide the same detailed individual, household, and housing variables found in the public-use version of the Census, but unlike the public-use data they provide information on the location of individual residences and workplaces at a very disaggregated level, down to the Census block level. Thus the restricted Census microdata allow us to identify the local neighborhood each individual inhabits, and to determine the characteristics of that neighborhood far more accurately than has been previously possible with such a large-scale data set.

Our study area consists of six contiguous counties in the San Francisco Bay Area: Alameda, Contra Costa, Marin, San Mateo, San Francisco, and Santa Clara. Though the framework we set out below has broad applicability for understanding segregation patterns, we focus on this area for three main reasons. First, it is reasonably self-contained. Examination of Bay Area commuting patterns in 1990 reveals that a very small proportion of commutes originating within these six counties ended up at work locations outside the area, and similarly a relatively small number of commutes to jobs within the six counties originated outside the area. Second, the area contains a racially diverse population, with significant numbers of Asian, Black, and Hispanic households. And third, the area is sizeable along a number of dimensions. The six counties include over 1,100 Census tracts, and almost 39,500 Census blocks, the smallest unit of aggregation in our data.[6] Our final sample consists of about 650,000 people in just under 244,000 households.

The Census provides a wealth of data on the individuals in the sample – their race, age, level of educational attainment, income, occupation (if working), language ability, marital status, and more. Throughout our analysis, we treat the household as the decision-making unit and characterize each

[6] Our sample consists of all households who filled out the long-form of the Census in 1990, approximately 1-in-7 households. In our sample, Census blocks contain an average of 6 households, while Census block groups – the next level of aggregation up - contain 92 households.

household's race as the race of the 'householder' – typically the household's primary earner. We assign households to one of four mutually exclusive categories of race/ethnicity: Hispanic, non-Hispanic Asian, non-Hispanic Black, and non-Hispanic White.[7] To ensure that our sample is representative of the overall Bay Area population, we employ the individual weights given in the Census. Accordingly, 12.3 percent of households are categorized as Asian, 8.8 percent as Black, 11.2 percent as Hispanic, and 67.7 percent of households as White.[8] The Census housing record provides other information on household characteristics, such as household size, family structure, number of children and languages spoken.

Using individual and household data linked to Census blocks, we have constructed a series of variables characterizing the neighborhood in which a household lives. We define a variety of neighborhoods based on conventional Census boundaries – the block, block group, tract, Public Use Microdata Area (PUMA) and county. In addition, as we know the latitude and longitude of the area center of each Census block, we define a succession of neighborhoods surrounding a given block that include all households in the sample in blocks within certain radii - half a mile, one mile, two miles etc. Using this approach, we can construct racial, education and income distributions based on the households in a given neighborhood surrounding each Census block. These provide the basis for our analysis of segregation. The full list of variables used in the analysis, along with means and standard deviations, is given in the Data Appendix.

[7] The task of characterizing a household's race/ethnicity gives rise to the issue of what to do with mixed race households. One solution would be to conduct the analysis at the level of the individual. Another solution would be to assign a household with, for instance, one White and one Hispanic individual a 0.5 measure for both categories and continue to keep the analysis at the household level, while a third option would be to use the characteristics of the household head to define the race/ethnic makeup of the household. We use this third approach and also omit the households that do not fit into one of these four primary racial categories (0.7 percent of all households). The results of our analysis are not sensitive to these decisions. Our final sample consists of the 243,350 households that fit into these four racial categories and live in a Census block group that contains at least one other household in our sample.

[8] The proportion of Whites is lower if we calculate the racial composition of the Bay Area based on all residents rather than just householders. The Census sample is highly representative of the Bay Area's population: If we calculate unweighted samples using the numbers of householders, 12.4 percent of households are characterized as Asian, 7.6 percent as Black, 10.9 percent as Hispanic, and 68.6 percent as White (and only 0.7 percent of households characterized as "Other"). These are very similar to the weighted proportions using Census 'person' weights.

II Patterns of Racial Segregation in the Bay Area

A. *Measurement Framework*

We begin our analysis by characterizing the patterns of racial segregation in the Bay Area. Given the assignment of households to one of the four primary race categories - Asian, Black, Hispanic, and White - we define dummy variables, r^i_j, that take the value one if household *i* is of race *j*, and zero otherwise. For a particular neighborhood definition, we calculate the fractions of households in each of the four racial categories that reside in the same neighborhood as a given household; let the upper-case notation R^i_k signify the fraction of households of race *k* in household *i*'s neighborhood. By averaging these neighborhood measures over all households of a given race, we construct measures of the average neighborhood racial composition for households of that race. Put another way, we construct measures of the average exposure, $E(r_j,R_k)$, of households of a race *j* to households of race *k*:

$$(1) \qquad E(r_j, R_k) = \frac{\sum_i r^i_j R^i_k}{\sum_i r^i_j}$$

An alternative and convenient way to construct these exposure rates is to run the following set of simple regressions. For each household *i*, regress R^i_k on the set of dummy variables r^i_j:

$$(2) \qquad R^i_k = \sum_j \gamma_{jk} r^i_j + \omega^i_k, \quad k \in \{A, B, H, W\},$$

where *k* ranges over the four race categories. The resulting parameters γ_{jk} are identically the average exposure of households of race *j* to race *k*, $E(r_j,R_k)$. This approach also provides a convenient way to distinguish the precision of these exposure rate measures – as the regression in equation (2) also provides standard errors for these measures.

A number of segregation measures are available,[9] and while no single measure is perfect, we choose to work with measures of segregation based on the exposure rates described above because

[9] Multiple measures of segregation have been used in the literature. (See Reardon and Firebaugh (2002) for a listing of multiple measures.) The measure most often used in sociology is the dissimilarity index (see Lieberson and Carter (1982) for example). Dissimilarity indices, which range between zero and one, provide information about the residential concentration of one race relative to others, specifically the share of one population that would need to move in order for the races in a region to be evenly distributed (see Cutler *et al.* (1999) for a definition). In contrast, the exposure rate measures used here simply returns the average rate of contact between people with specified sets of characteristics. The main benefit of using

exposure rates are easy to interpret and can be decomposed in a variety of meaningful ways. It is straightforward, for example, to calculate exposure rates for various subsets of households within each broad category (*e.g.* households of the same race but differing in their education levels), rates that must as a matter of necessity aggregate back up to the average exposure rate for the whole group. Unlike many segregation measures, exposure rates also allow us to examine the propensity of households of any pair of races to live together and to consider the factors that affect this propensity separately for different pairs of races. Thus we can see if households are clustering with specific households of other types rather than just examining own group sorting patterns.

Note that under the current approach, including a household's own race when constructing the neighborhood racial composition for that household can affect the measured exposure rates for our smaller neighborhood measures, for instance Census blocks rather than tracts. If, for example, a "neighborhood" always consisted of two households, then any Hispanic family would be in a "neighborhood" that was either 50 percent or 100 percent Hispanic. To avoid this problem, we define the racial makeup of a neighborhood to be the racial makeup of all *other* households in the neighborhood, and avoid including the individual household's own observation. It is important to point out that once this adjustment is made, any incorrect measurement of the neighborhood racial composition variables arising because of the small number of observations used to construct our smaller neighborhood measures does not bias the exposure rate measures.[10]

---

exposure rate measures arises when the analysis considers many household types (i.e., many categories of race/ethnicity or many other household characteristics). In this case, exposure rate measures provide information not only about the degree of clustering of households of a particular type, but also about the clustering of household each pair of household types.

Entropy measures are also used to measure segregation (see Massey and Denton (1989) for a description). Entropy measures summare the degree to which the racial distributions of neighborhoods within a region differ from the region's overall racial distribution, entropy being maximized for the region when the racial distributions at lower levels of aggregation are the same as that for the region overall (see, for instance, Harsman and Quigley (1995)). Finally, Borjas (1998) makes use of individual data, constructing a measure of segregation that takes the value one if the proportion of the individual's own ethnic group in the neighborhood is more than twice the proportion that would be expected under random assignment of individuals, an approach that loses information about the precise extent of local segregation.

[10] The intuition for this is easy to see in the context of the regression equations (2), as this is just a simple instance of white-noise measurement error in the dependent variable of this regression, which, unlike measurement error in the regressors, does not bias the parameter estimates.

It is possible to define a neighborhood and thus R^i_k in a number of ways. In the results that follow, we use the standard neighborhood measures given in the Census, rather than neighborhoods falling within given radii around each house.[11] These methods yield very similar results.

B. *Segregation Patterns*

Figure 1 provides information about the racial composition of Census block groups for the geographic core of our study area including San Francisco, Oakland, and Berkeley.[12] In the figure, block groups are shaded in distinct ways if they contain a majority of Asian, Black, or Hispanic households or more than 80 percent White households. Although Black households make up only 9 percent of the Bay Area population, the large number of Census block groups with a majority of Black households indicates a high degree of Black segregation. Census block groups with high concentrations of White households are clustered in Marin County and the more suburban areas of other counties while majority-Asian block groups are concentrated in San Francisco and Oakland. And although Hispanic households account for a higher proportion of the Bay Area population than Black households, there are far fewer Census block groups in which a majority of households are Hispanic.[13]

Table 1 provides the exposure rate measures described above calculated for Census block groups.[14] The table should be read as follows: consider the measured exposure rates of the typical Asian household at the Census block group level shown in the top panel of the table. Reading across the first row, these measures imply that Asian households live in Census block groups that have on average 23 percent Asian households, 8 percent Black, 12 percent Hispanic, and 57 percent White households. Comparing these numbers to the racial distribution of the Bay Area as a whole, given in

[11] We considered both methods of defining neighborhoods, as the first corresponds to the approach most commonly used in the literature and the second might provide a better approximation to a household's neighborhood in certain cases. The usual method of looking for segregation patterns across well-defined geographic units like Census tracts might give misleading results if, for example, households are sorted within the tract so that they match up with the households in neighboring tracts. However, analyses examining neighborhoods defined as observations falling within 0.25, 0.5 and 1-mile radii of a Census block produced results similar to those for Census block groups (.25 miles) and tracts (.5 and 1-mile radii).
[12] Figure 1 is derived from information in the public-use Census data set.
[13] It is worth noting that if a more aggregate definition of neighborhood is used the percent of majority one race neighborhoods declines substantially. At the PUMA level there are only four PUMAs that are mostly segregated, an area of Marin and the outlying areas of Contra Costa county are over 80% White and a PUMA in Alameda County is primarily Black.
[14] The regression results underlying these exposure rates and their calculated standard errors can be found in Appendix Table 1. As one would expect with nearly a quarter of a million observations, these exposure rate measures are estimated very precisely.

the row labeled "Overall" - 12 percent Asian, 9 percent Black, 11 percent Hispanic, and 68 percent White - it is apparent that the typical Asian household lives in a Census block group with approximately twice the fraction of Asian households as would be found if they were uniformly distributed across the Bay Area. In this case, the additional fraction of Asian households in Census block groups in which Asian households reside is almost exactly offset by a reduction in the fraction of White households in these neighborhoods,[15] with Black and Hispanic households being found in roughly the same proportions as their overall proportions for the Bay Area.

Examining the exposure measures for each race at the Census block group level, a clear pattern emerges, with households of each race residing with households from the same race in proportions significantly higher than their proportions for the Bay Area as a whole. Not surprisingly given the geographic concentration shown in Figure 1, the most striking example of such 'over-exposure' of households to other households of the same race occurs for Black households. On average, the typical Black household lives in a Census block group that has almost 5 times the fraction of Black households as the whole Bay Area and over 8 times the average fraction of Black households as are found in the neighborhoods inhabited by White households. The pattern for Hispanic households is similar to that for Asian households, with the typical Hispanic household living in a block group that has almost twice the proportion of Hispanic households as the Bay Area as a whole, slightly higher proportions of Asian and Black households, and a lower proportion of White households than are found in the Bay Area as a whole (56.2 percent vs. 67.7 percent). Consistent with the previous patterns, White households on average live in block groups with a lower proportion of other races than would be found if all racial groups were evenly spread across block groups. However, the 'under-representation' of Black households (5 percent vs. 9 percent) in neighborhoods in which White households reside is more sizeable than that of Asian (10 percent vs. 12 percent) and Hispanic (9 percent vs. 11 percent) households.

We present exposure rates at five levels of aggregation - county, PUMA, tract, block group, and block - in Appendix Table 2. Examining these exposure rates, it is clear that the exposure of households to other households of the same race increases as the size of the geographic unit under

[15] It is worth noting that other segregation measures such as dissimilarity indices would miss the fact that the increased exposure of typical Asian, Black, and Hispanic households to other households of the same race is almost completely offset by a decreased exposure to White households. We also find that black households live with proportionately fewer Asian households and more Hispanic households, although these differences are dwarved by the decline in white households.

consideration declines. While this general trend is not surprising, the extent to which these measures differ for PUMAs, which contain approximately 50,000 households, and smaller Census areas such as block groups (around 500 households) and blocks (around 50 households) is significant. The exposure rate measures in Appendix Table 2 imply, for example, that an analysis of segregation at the PUMA level, which is the smallest geographic unit specified in the public-use Census microdata, would significantly understate the fraction of immediate neighbors who are of the same race. This points to the importance of using the restricted data for the type of household-level analysis conducted in the current paper.

II Exploring the Mechanisms Underlying Segregation – An Illustration: Education

Having characterized the general patterns of racial segregation in the Bay Area, we now turn to the main analysis of the paper - examining the extent to which the correlation of race with other household attributes can explain the segregation of each race. Previous studies that have attempted to examine this question have employed a range of empirical strategies (see Massey and Denton (1993), (1998), and Harsman and Quigley (1995)), but have often been restricted by the structure of available data. In particular, for studies based on relatively small geographic areas, researchers have known only the marginal distributions of race, education, income, and other household attributes. In the current analysis, we seek to exploit the richness of our restricted Census data, in particular the fact that we know the *joint* distribution of household characteristics at very low levels of geographic aggregation. That is we can examine *for a household* the characteristics of neighboring households of each *race separately*

To this end, we develop an empirical strategy that builds on the approach taken by Borjas (1998). Here, we require two conditions to hold in order to conclude that a particular household characteristic to explain patterns of racial sorting. *First, the distribution of this household characteristic must differ significantly across race.* If, for example, the distribution of educational attainment were the same for all races, it seems reasonable to conclude that this factor would have no ability to explain the observed pattern of racial segregation. *Second, the attribute in question must affect the typical racial composition of the neighborhoods in which households of a given race live.* If Hispanic households, for instance, were exposed to the same fraction of other Hispanic households regardless of income, it seems reasonable to conclude that differences in income between Hispanic households and the other households in the Bay Area could not explain the segregation of Hispanic households.

To determine the household attributes that satisfy the first condition described above, Table 2 summarizes a series of household attributes by race. It is immediately apparent that households of different races differ along many other dimensions, including education, income and wealth, family structure, language(s) spoken, and citizenship. The first five rows show the distribution of education attainment across households of different races. For Asian households, the distribution of educational attainment is more dispersed than the overall sample; that is, more Asian household heads have not completed high school (19 percent vs. 16 percent) or have an advanced degree (16 percent vs. 14 percent) than in the sample as a whole.[16] In contrast, Black and Hispanic households have less education on average than the sample overall and White households are more likely to be headed by someone with a bachelor's degree or higher (49 percent for White households vs. 43 percent overall).

Examining the remaining rows, it is clear that Hispanic and Asian households are more likely to speak a language other than English in their homes, more likely to be immigrants and more likely to have recently arrived in the United States than Black and White households. And Black households have lower income, are more likely to receive public assistance and much less likely to have dividend or capital gains income than households of other races. Thus a number of household attributes have the potential to explain the segregation of households of each race.

A. *An Illustration of a Counterfactual Simulation: Education*

We begin our analysis by considering the extent to which racial segregation can be explained by differences in educational attainment. This example not only demonstrates the two measurement approaches we use, but also clarifies the basic assumptions underlying the empirical analyses conducted in the paper. It should be noted, of course, that this preliminary example controls only for

---

[16] This dispersion is caused mainly by dispersion across Asians of different nationalities, with certain groups having lower educational attainment than others (*e.g.* Vietnamese compared to Japanese). Carrying out separate analyses for the different Asian nationalities, we find that the patterns found for Asians generally are repeated, by-and-large. Confidentiality requirements restrict the reporting of these disaggregated findings.

educational attainment; adding other household characteristics will tend to diminish the amount of segregation explained by racial differences in education[17].

Table 2 makes clear that the distribution educational attainment varies significantly across race. Not surprisingly, education also plays an important role in the sorting process as shown in Table 3, which characterizes the stratification of households in the Bay Area across Census block groups on the basis of education. We divide educational attainment into five categories – less than high school degree (e_1), high school degree (e_2), some college (e_3), bachelor's degree (e_4), and advanced degree (e_5) – and run the following set of regressions, similar to those in equation (2) for race/ethnicity:

$$(3) \qquad E_k^i = \sum_{j=1}^{5} \alpha_{jk} e_j^i + \upsilon_k^i$$

Here, analogous to the race regressions used to generate Table 1, α_{jk} represents the exposure of households of education level j to households of education level k.

The results in Table 3 reveal a significant amount of stratification on the basis of education. While approximately 15 percent of households in our sample have less than a high school degree and 15 percent have an advanced degree, households at different ends of the educational attainment spectrum typically live in quite different neighborhoods, based on the education levels of their inhabitants. A household with less than a high school degree lives in a block group with 26 percent of households with less than a high school degree and only 8 percent with more than a BA on average, while a household headed by someone with an advanced degree typically resides in a neighborhood with 9 percent of households with less than a high school degree and 23 percent with an advanced degree. As with race, households with a given level of education attainment are more likely to live with other households with the same level of educational attainment than would be predicted if households were evenly distributed throughout the Bay Area.

The combination of the sorting of households on the basis of education and the significant differences in education across races (described in Table 2) suggests that differences in education may explain a substantial amount of racial segregation. Table 4 shows exposure rates for each race by the educational attainment of the head of household. Note that for each race, as a household's education

[17] We will also examine the impact of changing the distribution of educational attainment using another simulation methodology. This second example will take into account that the clustering found in the overall neighborhood (left hand side percentages) will change as the individual household education distribution is reallocated.

increases, so the percentage of White households in the neighborhood in which they live also increases monotonically. For Asian, Black, and Hispanic households, this increasing exposure to White households coincides with a decreasing exposure to households of the same race. In addition, for almost all groups the exposure to Black and Hispanic households declines as educational attainment increases.[18] Thus, while a typical Black household headed by someone who is a high school dropout lives in a block group in which 53 percent of the households are Black, this level is halved for a household headed by someone with an advanced degree.[19]

In order to understand the role of education in driving racial segregation, we seek to determine how the pattern of racial segregation would change if across-race differences in education were eliminated - in other words, if each race had the empirical distribution of education observed in the population of the Bay Area as a whole. To conduct this type of counterfactual, it is necessary to make an assumption about the features of the observed pattern of household sorting that would be unaffected by such an adjustment to the education distribution of each race. In the analysis that follows, we consider two alternative assumptions concerning the *primitives* of the observed pattern of household sorting. While we do not model the sorting process directly, these assumptions allow us to exploit the richness of our data to learn about the driving forces behind segregation in a reasonable and straightforward manner. We discuss the relative merits of these alternative assumptions after applying each to the example of educational attainment.

As a first approach, we take the exposure rates of Table 4 to be primitives and calculate the new average exposure rates that result from shifting the education distribution of each race to the mean. In this way, we assume, for example, that the exposure of highly educated White households to Hispanic households would be unaffected by a change in the education distribution of Hispanic households.

The results of this exercise are shown in Table 5[20]. The first column of Table 5 presents the exposure rates of households to those of the same race, drawn from Table 4. Columns (2) and (3) in

---

[18] The one exception is that for black households an increase from ho high school diploma to having a high school diploma slightly increases the average percentage of Hispanic households in the neighborhood.

[19] It is interesting to note that, for Black households, increasing education *increases* the percentage of Asian households in the neighborhood, while the percentage of Hispanic households declines. As Asian households become more educated, they typically live in communities with fewer Black and Hispanic households.

[20] We carry out a similar exercise for a full set of household characteristics later in the paper. This calculations carried out here are similar in spirit to the decompositions done by Juhn, Murphy and Pierce

each panel present the educational attainment distribution of each race and that of the overall sample, respectively. The fourth column then uses the actual education distribution for each race to calculate the overall own-race exposure rate, given at the bottom of the column in each of the four panels. This number is the same as the own-race exposure rate given in Table 1, as expected when using the actual education distribution to provide weights. The fifth column calculates the own-race exposure rate under the counterfactual that the particular race shares the education distribution of the overall sample. Comparing the numbers at the bottom of columns (4) and (5) shows how much exposure rates would change as a result of changing the education distribution of each race. Column (6) reports the percentage reduction in the 'over-exposure' of each race to other households of the same race, where 'over-exposure' is defined relative to the fraction of households of each race in the full Bay Area sample. That is, Column 6 calculates reduction in the level of over-exposure to households of the same race when differences in education are controlled for divided by the level of over-exposure found on average for households of a given race. (For example for Hispanic households this equals (.221-.181)/(.221-.112)=.37 or a 37 percent reduction in the over-exposure rate)

As the numbers in Table 5 indicate, the impact of changing the education distribution varies by race. For Hispanic households, using the mean education distribution would reduce the average exposure of Hispanic households to other Hispanic households from 22 to 18 percent, almost a 37 percent reduction in the 'over-exposure' of Hispanic households to other Hispanic households. This suggests that over one-third of the segregation of Hispanic households can be explained by the fact that Hispanic households have relatively low levels of education. Likewise, education can explain about eight percent of the segregation of Black households, seven percent of White segregation, and one percent of Asian segregation.

B. *An Alternative Assumption about the Primitives of the Sorting Process*

It is likely, of course, that the own-race exposure rates shown in the Table 4 and the first column of Table 5 would be affected by a change in the underlying education distribution of each race. If, for example, Hispanic households with low levels of education have a strong tendency to live with one another, a decrease in the fraction of Hispanic households with a high school degree from 39 percent to the sample mean of 16 percent would likely reduce the average own-race exposure rate of

---

(1993) to examine changes in wage inequality and by Reed (1999) to study causes of income inequality in California.

poorly educated Hispanic households. At the same time, an increase in the mean education level of Hispanic households would likely increase the exposure of highly educated Hispanic households to Hispanic households in general.

Based on these considerations, we make an alternative assumption concerning the primitives of the sorting process, utilizing measures of the exposure of households in each race-education category to households in every other race-education category. As an alternative to the fixed exposure rate assumption used above, we treat the propensity to live with households in each race-education category relative to the fraction of households in that category in the full sample as the primitive of the underlying sorting process. We label this relative exposure measure the *intensity of exposure* to households in each race-education category. Thus the exposure of highly educated White households to Hispanic households, for example, is allowed to increase with an upward shift in the Hispanic education distribution, provided highly educated White households have a greater intensity of exposure to highly educated versus poorly educated Hispanic households. Having calculated the new exposure rates implied by the shifts in the education distribution, we repeat the analysis from above using these adjusted exposure rates.[21] In this counterfactual we again have the attractive feature of the overall distribution of households with a given level of education or race adding up to the actual number of households. That is the increase in educational attainment found for Hispanic households is exactly offset by a shift in the educational attainment of other households.

Because the full set of interactions would be too cumbersome to report (20 race-education categories leads to 400 cells), Table 6 shows the results for the own-race exposure of Hispanic households to illustrate the procedure. The upper panel in Table 6 shows the average fraction of Hispanic households in each education category that reside in the neighborhood in which Hispanic households with the education level listed in the row heading reside. For example, the first row provides the average exposure of Hispanic households without a High School diploma to Hispanic households in each education category. As the table shows, an average of 17 percent of the neighbors of Hispanic households without a High School diploma are also Hispanic households without a High School Diploma while an average of only half of one percent are Hispanic households with a post-graduate degree. The next four rows show the same kind of distributional information for Hispanic

[21] We should emphasize that this is not the only alternative to the 'fixed exposure rates' approach used above. However, it does provide a systematic way of carrying out more flexible counterfactuals, thereby providing a useful comparison to the counterfactuals based on fixed exposure rates.

households with higher education levels, while the final row in this upper panel shows, for comparison, the fraction of the Bay Area's population accounted for by Hispanic households in each education category. The right-most columns of the upper panel of Table 6 simply repeat the calculations of Table 5 for the sake of comparison.

The middle panel in Table 6 then calculates the intensity of exposure for Hispanic households with a given level of education to Hispanic households in each education category. The intensity of exposure for a given education pair is just the ratio of the average fraction of Hispanic households of a given education level in the neighborhood to the overall fraction of Hispanic households with that education level in the Bay Area. Thus, Hispanic households headed by householders without a High School Diploma are typically exposed to almost four times as many households of the same type than would be expected in the overall sample (16.5 percent vs. 4.4 percent). The fact that almost all of the figures in this middle panel are greater than one implies that Hispanic households are exposed to a greater fraction of Hispanic households in almost every education category than the fraction of Hispanic households in that education category in the Bay Area as a whole. Moreover, the greatest intensities of exposure in the table describe the propensity of Hispanic households with low levels of education to live together.

The bottom panel in Table 6 uses the intensity of exposure measures from the middle panel to calculate new exposure rates under the counterfactual that Hispanic households had the education distribution of the Bay Area as a whole; and recall that the intensity of exposure measures are taken as the primitives of the sorting process in this counterfactual. In this case, a typical Hispanic household with less than a High School Diploma is predicted to live in a neighborhood in which 6.8 percent of households are Hispanic households with less than a High School Diploma. This number is calculated by taking the adjusted fraction of Hispanic households in the Bay Area with less than a High School Diploma – 1.8 percent – and scaling it up by the fixed intensity of exposure rate of 3.8 for that race education pair.

The sixth column of this bottom panel shows how the overall own-race exposure of Hispanic households in each education category changes as a result of treating the intensity of exposure measures as primitives. As the figures in this column illustrate, treating the intensity of exposure measures as primitives greatly reduces the exposure of Hispanic households in the lowest education categories to other Hispanic households. Put another way, because Hispanic households with low levels of education have such strong intensities of exposure to other poorly educated Hispanic households, the upward shift in the education distribution dramatically reduces the overall own-race exposure of these households. At the same time, because Hispanic households with a bachelor's degree, for example, tend to be exposed in roughly the same intensity to Hispanic households in all education categories, the overall own-race exposure of these households changes very little.

The rightmost columns of the bottom panel of Table 6 calculate the average exposure of Hispanic households to other Hispanic households using the new exposure rates and new weights based on the education distribution of the full population of the Bay Area. The predicted reduction in the 'over-exposure' of Hispanic households to one another using the intensity of exposure measures as primitives is now 55 percent compared with 36 percent when the exposure rates themselves were used as primitives (in previous sub-section).

C. *Comparing the Results of the Alternative Counterfactuals*

In the light of these findings, the results of the first counterfactual that treated the exposure rates of Table 4 as primitives understates the impact of eliminating educational differences across race in reducing the segregation of Hispanic households relative to the counterfactual that treated intensities of exposure as primitives. This turns out to be a robust feature of our analysis, holding for the impact of each household characteristic on the segregation of each race. While we provide more evidence for the full set of household characteristics at the end of the next section, the calculations in Table 6 provide a clear understanding as to why this occurs when examining the impact of education on Hispanic segregation. The greater reduction in Hispanic own-race exposure when the intensity of exposure measures are treated as the primitives results from two features of the data. First, Hispanic households tend to have lower levels of education than the Bay Area population as a whole. Second, the intensity of exposure measures are greatest for the exposure of Hispanic households with low levels of education to one another. So when the distribution of Hispanic education is increased in the second counterfactual, weight is shifted away from the portions of the intensity of exposure matrix with the largest intensities. This leads to the dramatic reduction in the exposure of poorly educated Hispanic households to Hispanic households in general discussed above, thereby leading to a greater reduction in the average own-race exposure of Hispanic households, relative to our initial counterfactual. We provide further evidence concerning the conservative nature of the counterfactuals that treat exposure rates as primitives at the end of Section IV. Again, note that in the previous examples educational attainment differences were the only household characteristics allowed to impact the distribution of races across neighborhoods and may have been picking up the impact of other household characteristics that are correlated with education. For example, if most Hispanic households with less than a high school education are immigrant households the education effects may be picking up an immigrant effect.

IV Exploring the Mechanisms Underlying Segregation – The Full Analysis

We now extend the analysis to examine the ability of a full set of household attributes to explain observed segregation patterns. To measure how household characteristics affect the exposure of households of race j to households of race k, we include interactions of household attributes and household race in the regressions developed in equation (2):

$$R_k^i = \sum_m \sum_j \gamma_{jkm} r_j^i x_m^i + v_k^i, \quad k \in \{A, B, H, W\}. \tag{4}$$

Here, each variable x_m represents a household attribute and each parameter, γ_{jkm}, describes how attribute x_m affects the exposure of households of race j to race k.

By multiplying each of the resulting parameters γ_{jkm} by the mean of each household attribute for race j, $\overline{x}_{jm}$, and summing over the included attributes, we reproduce the average exposure of households of race j to households of race k:

$$E(r_j, R_k) = \sum_m \gamma_{jkm} \overline{x}_{jm} = \gamma_{jk} \tag{5}$$

Substituting instead the mean of each household attribute from the full sample, $\overline{x}_m$, we calculate what we term 'the average exposure of households of race j to households of race k conditional on the set of attributes X,' labeled $E(r_j, R_k \mid X)$:

$$E(r_j, R_k \mid X) = \sum_m \gamma_{jkm} \overline{x}_m \tag{6}$$

By comparing $E(r_j, R_k \mid X)$ to $E(r_j, R_k)$, we calculate the impact of reducing across-race differences in all of the included household attributes X on the exposure of households of race j to households of race k. Having estimated equation (4) with a full set of interactions, we calculate the *marginal* impact of a particular household attribute on the exposure of race j to race k by replacing $\bar{x}_j$ with $\bar{x}$ for only that attribute.

A. *Predicting Exposure to Households of the Same Race*

Because the four mutually exclusive categories of household race are interacted with each household attribute in the regressions shown in equation (4), it is possible to produce the same parameters by stratifying the sample by race and running separate regressions for each race. The resulting parameter estimates describe how each household attribute affects the propensity of households of the race by which the sample is stratified to live with households of the race that constitutes the dependent variable. In order to keep the results tractable, we report only four of the full sixteen regressions in Table 7 - those that describe how household attributes affect the propensity of households of each race to segregate from or live with households of the same race.

The first rows of Table 7 show the marginal impact of educational attainment on the propensity of households of each race to live with others of the same race.[22] These results show the same patterns as Table 4, but not surprisingly, the magnitudes are significantly reduced. For example, at the margin, Black households with less than a high school degree live in neighborhoods with 12 percentage points more Black households than Black households with an advanced degree. This compares to the 27-percentage point difference shown in Table 4. The next set of rows show the impact of household income on racial stratification. As with education, increases in income lead to more segregation on the part of White households and less on the part of households of other races. Likewise, we find that the impact of income is largest for Black households. The source of income, in addition to the magnitude, also affects the propensity of households of each race to live with other households of the same race. Black and Hispanic households with capital income tend to live with fewer households of the same race, while Hispanic and especially Black households with public assistance income are more likely to be segregated. Not surprisingly, we also find that speaking a language other than English increases the level of segregation for Asian and Hispanic households, as

[22] Exposure rates can be recovered from these estimates by adding coefficients for households of a given race and given characteristics to the race-specific constants at the bottom of each column.

does answering that the household head speaks only some English or no English. There is also an increase in the segregation of households of all races who have recently moved to the US and of all races other than Black households that are naturalized or not US citizens, especially Asian households.

B. *Explaining Segregation – The Full Set of Household Characteristics*

Table 8 presents the results of our first counterfactual calculations that treat the parameters of the regressions shown in Table 7 as primitives. As in the first counterfactual exercise of the previous section that focused only on education, these counterfactuals do not account for the fact that the exposure rates implied by the regressions in Table 7 might themselves adjust as the underlying characteristics of each race change. As in the previous section, we consider below an alternative assumption that treats the intensity of exposure as a primitive. The top panel of Table 8 gives, for each race, the percentage of racial segregation that can be explained by non-racial household characteristics. The first set of rows presents information first shown in Table 1, that is the overall distribution of each racial group and the over-exposure of the average household of each race to other households of the same race. The next set of rows presents the over-exposure rate that would occur if there were no differences in household characteristics across each racial group that is it estimates the percent of households predicted to live in a neighborhood of the same race using the regression estimates and the overall sample means. Rows 5 and 6 then relate the decline in exposure rates due to differences in household characteristics to that originally found. As the last row in this panel indicates, differences in non-racial household attributes together explain approximately 93 percent of segregation for Hispanic households, 53 percent for Asian households, 32 percent for White households, and 30 percent for Black households. Note that although an equal amount of the over-exposure rates for Black and White households occurs the relative amount of over-exposure was much higher for Black households. Note this exercise is similar to the one we carried out for educational attainment in Table 5.

To understand which household attributes drive the segregation of each race, we decompose the overall percentages reported in the lower panel of Table 8. This lower panel shows the *marginal* effects of five different sets of attributes: educational attainment, income, language, citizenship, and household demographics. In each case, we calculate exposure rates when the distribution of a particular set of attributes for each race is replaced by the mean distribution of that set of households in the overall sample using the approach described above and then list the amount of the decline in over-exposure of households related to the given attribute. We discuss the findings for each race in turn.

For Asian households, the primary driver of segregation relates to language, which alone can account for almost 40 percent of the 'over-exposure' of Asian households to other Asian households. Interestingly, much of this effect derives from whether another language is spoken rather than how well English is spoken in the household. Since 75% of Asian Households speak an Asian language the results imply that Asian households that do not know another language resemble the overall population. Factors related to immigration status and citizenship explain another 8.5 percent of Asian segregation. Income, education, and family structure have little to no explanatory power.

Lower levels of income, as well as the higher probability of drawing public assistance and lower probability of having capital income, increase the segregation of Black households, explaining over 14 percent of the 'over-exposure' of Black households to other Black households. Differences in education and factors related to immigration and citizenship explain another 11 percent of Black segregation, but family structure variables explain very little.

For Hispanic households, almost every included set of household characteristics has some ability to explain Hispanic segregation. As in the case of Asian segregation, more than 30 percent of the residential concentration of Hispanic households can be explained by language differences, with much of this difference coming from speaking Spanish in the house. Lower than average levels of education and income explain another 19 and 10 percent of Hispanic segregation respectively and family structure – in particular, larger household sizes – explains another 14 percent. Notably, factors related to citizenship and immigration explain none of the observed segregation of Hispanic households on the margin. Combined with the similar finding for the relationship between language and immigration for Asian households, these results suggest that households who do not speak another language show little taste for living in neighborhoods with a larger concentration of other households of the same races. Alternately, if a non-immigrant family chooses to speak another language in the home this is an indication that they also prefer to live in neighborhoods with a higher concentration of other families of the same racial or ethnic background, rather than clustering being caused by an inability on non-English speakers or new immigrants being limited to specific neighborhoods. This preference could be driven by stores or other characteristics of these neighborhoods (ie preferences for stores that carry specific foods) rather than a preference to live with other like families.

The segregation of White households is driven by a variety of factors. The fact that White households have higher than average levels of income and education combined with the fact that White segregation increases with increasing levels of these characteristics implies that a portion of the over-exposure of White households to other White households can be explained by these factors – around 12 percent. Language differences can also account for about 15 percent of White segregation, while immigration status, citizenship, and family structure have almost no explanatory power. The language difference information may reflect that someone else in the household is of another racial or ethnic group.

While the inclusion of additional household attributes could further reduce the unexplained portion of racial segregation, we believe that the analysis in this section includes the household attributes that are most relevant to explaining a significant fraction of racial segregation. A number of potential explanations arise for this portion of segregation that cannot be explained by household characteristics. Households of different races may reside in different neighborhoods, for example, because they systematically demand different physical features of their house or neighborhood or because they have different preferences for the race or other characteristics of their neighbors or because they desire different levels of amenities affiliated with household location like quality of local schools or proximity to green space or highways. The role these neighborhood qualities play in determining household location preferences is explored in Bayer *et al.* (2002).

To the extent that the unexplained portion of the segregation of each race is directly related to race itself, it is also important to emphasize that our analysis provides no indication of the root cause of this portion of segregation. The segregation of Black households related to race itself could arise, for example, because of the preferences of Black households to live together, the preferences of Asian, Hispanic, or White households to live with others of the same race, the preferences of Asian, Hispanic, or White households to avoid Black households, or systematic differences in demand for housing and other neighborhood amenities across race, among other explanations. Our analysis provides no evidence that can distinguish these and other alternative explanations for the unexplained portion of racial segregation. Again, we should also note that our analysis does not estimate the role race plays in the attainment of different household characteristics. That is, if discrimination leads to lower income and that leads to sorting of households we do not measure this indirect role as unexplained segregation.

C. *Treating Intensity of Exposure as a Primitive of the Sorting Process*

As in Section III, we again would like to consider an alternative set of counterfactuals that treat the intensity of exposure measures as the primitives of the sorting process. As the education example makes clear however, this type of counterfactual requires exposure rate measures for each distinct category of race and household characteristics interacted with every other distinct category. As the number and type of categories increases, this approach quickly exceeds the capacity of our data, despite the fact that we have almost a quarter of a million observations. Creating separate cells for all of the interactions included in the regressions of Table 7, for example, would require almost one billion distinct cells. In conducting the counterfactuals that treat the intensity of exposure measures as primitives, then, we focus on the effects of variables that are likely to have the greatest influence and consider a number of different groupings of household characteristic categories such that the total number of distinct cells is limited to 4096 (64 distinct race-household characteristic categories).

Table 9 presents the results from this exercise. For each distinct grouping, we also report analogous results based on counterfactuals that treat exposure rates as primitives, reported in the 'Fixed Exposure Rates' rows. The first panel of Table 9 simply repeats the education results from Section III and includes the five categories of education used in that analysis. The second panel creates twelve distinct categories of household characteristics (2 education categories x 3 income categories x 2 language categories). As in the previous education example, the counterfactuals that treat intensity of exposure as a primitive 'explain' a greater percentage of the segregation of each race (measured again here as the percentage reduction in own-race 'over-exposure' relative to the sample mean) than the counterfactuals that treat exposure rates as primitives. This general finding holds consistently in every alternative grouping that we have ever tried. The remaining panels of Table 9 consider additional household characteristics such as immigrant status and public assistance income in the formation of distinct categories of household characteristics. In all cases, the two counterfactuals produce a similar pattern of results with the fixed intensity of exposure counterfactuals increasing the explanatory power by an average of about 60 percent. In the light of these results, we conclude that the counterfactuals described in Table 8 that treat exposure rates as primitives and use the full set of characteristics reported almost certainly *underestimate* the amount of sorting explained by these household characteristics. At the same time, the analysis of Table 9 (especially the final panel) confirms our general findings in the first set of counterfactuals, namely that these other household characteristics explain the vast majority of Hispanic and to a lesser extent Asian segregation, while leaving much of the segregation of Black and White households unexplained.

This points to a direct trade-off between the two types of counterfactuals described in our analysis. While the calculations that use exposure rates as primitives almost certainly understate the ability of household characteristics to explain racial segregation, this approach allows us to simultaneously control for a wide range of household characteristics in the analysis. And while the calculations that use intensity of exposure measures as primitives are likely more appropriate counterfactuals, the data requirements quickly grow too large. In light of these limitations, we focus attention primarily on the former set of results, noting that the explanatory power of the included household variables is likely to be significantly but not overwhelmingly greater.

V Conclusion

Using unique and previously untapped micro-data drawn from the restricted-access version of the 1990 Census, the main goal of this paper has been to provide new evidence on an old question – to what extent can the correlation of race with other sociodemographic characteristics explain the observed degree of residential segregation on the basis of race? Though our analysis focused on the San Francisco Bay Area, the method has broader applicability, providing a clean way of both describing and decomposing patterns of neighborhood segregation, and of exploring relevant counterfactuals. Future work could certainly extend this analysis to a more nationally representative sample of metropolitan areas.

In line with the previous literature, our results indicate that segregation patterns vary markedly by race, though there is a tendency for households of a given race to cluster disproportionately with households of the same race. The extent of this clustering depends to a considerable degree on the definition of neighborhood used and we find that a substantial amount of segregation is missed when segregation patterns are studied at the county, PUMA, or even tract level.

In direct contrast to the previous literature, however, our findings indicate that household attributes, including education, income, language, and immigration status, can collectively explain almost 95 percent of the segregation for Hispanic households, over 50 percent for Asian households, and approximately 30 percent for White and Black households. For Hispanic households, racial segregation appears to be primarily a by-product of the sorting that occurs in any metropolitan area on the basis of education, income, language and other household attributes. In contrast, the results suggest that race itself directly contributes to the segregation of Black and White households. The results also provide a great deal of information about how a wide set of household characteristics affect the segregation patterns of households of each race, with a different set of household characteristics serving as the primary driver of the segregation of households of each race

In drawing attention to the importance of a variety of underlying factors driving the segregation of each race, our analysis provides the type of evidence that should guide policy aimed at reducing racial segregation. Again, given that speaking a language other than English serves as a more important factor in explaining Asian and Hispanic residential segregation than does language ability, this suggests that language instruction in schools may not change housing patterns.[23] The amount of segregation driven by differences in educational attainment indicates that policies aimed at improving the educational attainment of Hispanic and Black students are likely to have the indirect effect of reducing racial segregation. Segregation of Black and Hispanic households attributable to income (level and source) is more troubling, and may have been exacerbated by existing public policy. The fact that housing assistance has historically been administered to families living in large, somewhat isolated, housing projects is likely to have contributed to observed racial segregation patterns. New federal and state programs adopted during the past decade to replace large housing projects with smaller mixed-income properties and vouchers giving families more choice are less likely to reinforce segregation patterns.

[23] It is possible that speaking a language other than English may provide evidence of a preference on the part of certain Asian and Hispanic households to live in more segregated neighborhoods. This may have more to do with services found in such neighborhoods than discrimination or lack of opportunities for these households to live elsewhere.

References

Bailey, M.J., (1966), "Effects of Race and Other Socioeconomic Factors on the Values of Single-Family Homes," *Land Economics*, 42, 215-220.

Bajari, Patrick, and Matthew Kahn (2001), "Why Do Blacks Live in Cities and Whites Live in Suburbs?" unpublished manuscript, Stanford University.

Bayer, Patrick, Robert McMillan, and Kim Rueben, (2002), "The Causes and Consequences of Residential Segregation: An Equilibrium Analysis of Neighborhood Sorting," mimeo, Yale University.

Borjas, George J., (1995), "Ethnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review*, 85(3): 365-390.

Borjas, George J., (1998), "To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation," *Journal of Urban Economics*, 44: 228-253

Chambers, Daniel N, (1992), "The Racial Housing Price Differential and Racially Transitional Neighborhood," *Journal of Urban Economics*, 32, 214-232.

Cutler, David M. and Edward L. Glaeser, (1997), "Are Ghettos Good or Bad?" *Quarterly Journal of Economics*, August: 826-72.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor (1999), "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107(3): 455-505.

Farley, Reynolds and William H. Frey. "Changes in the Segregation of Whites from Blacks During the 1980s: Small Steps to a More Integrated Society." *American Sociological Review* 59:23-45

Foillain, James R. and Stephen Malpezzi, (1981), "Another Look at Racial Difference in Housing Prices," *Urban Studies*, 18: 195-203.

Frey, William H. and Reynolds Farley. "Latino, Asian, and Black Segregation in U.S. Metropolitan Areas." *Demography* 33(1):35-50

Gabriel, Stuart and Stuart Rosenthal, (1989), "Household Location and Race: Estimates of a Multinomial Logit Model," *Review of Economics and Statistics*, 71: 240-249.

Harsman, Bjorn and John M. Quigley, (1995) "The Spatial Segregation of Ethnic and Demographic Groups: Comparative Evidence from Stockholm and San Francisco," *Journal of Urban Economics*, 37: 1-16.

Juhn, C., K. Murphy, and B. Pierce (1993), "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy,*101:410-442.

Kiel, Katherine and Jeffrey Zabel, (1996), "Housing Price Differentials in U.S. Cities: Household and Neighborhood Race Effects," Journal of Housing Economics, 5: 143-165.

King, Thomas and Peter Mieszkowski, (1973), "Racial Discrimination, Segregation, and the Price of Housing," Journal of Political Economy, 81: 590-606.

Lieberson, Stanley, and Donna K. Carter, (1982), "Temporal Changes and Urban Differences in Segregation: A Reconsideration," *American Journal of Sociology* 88: 296-328.

Massey, Douglas S., and Nancy A. Denton, (1987), “Trends in the Residential Segregation of Blacks, Hispanics, and Asians,” *American Sociological Review*, 52: 802-825.

Massey, Douglas S., and Nancy A. Denton, (1988), “Suburbanization and Segregation in U.S> Metropolitan Areas,” *American Journal of Sociology*, 94(3): 529-626.

Massey, Douglas S., and Nancy A. Denton, (1989), “Hypersegregation in United States Metropolitan Areas – Black and Hispanic Segregation along Five Dimensions,” *Demography*, 26: 373-91.

Massey, Douglas S., and Nancy A. Denton, (1993), *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Miller, V. and John M. Quigley, (1990), “Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area,” *Urban Studies*, 27: 3-21.

Reardon, Sean and Glenn Firebaugh (forthcoming), “Measures of Multigroup Segregation,” *Sociological Methodology*

Reed, Deborah (1999), *California’s Rising Income Inequality: Causes and Concerns.* San Francisco, CA: Public Policy Institute of California.

Schafer, Robert, (1979), “Racial Discrimination in the Boston Housing Market,” *Journal of Urban Economics*, 6: 1176-1196.

Schelling, Thomas C., (1969), “Models of Segregation,” *American Economic Review*, 59(2): 488-93.

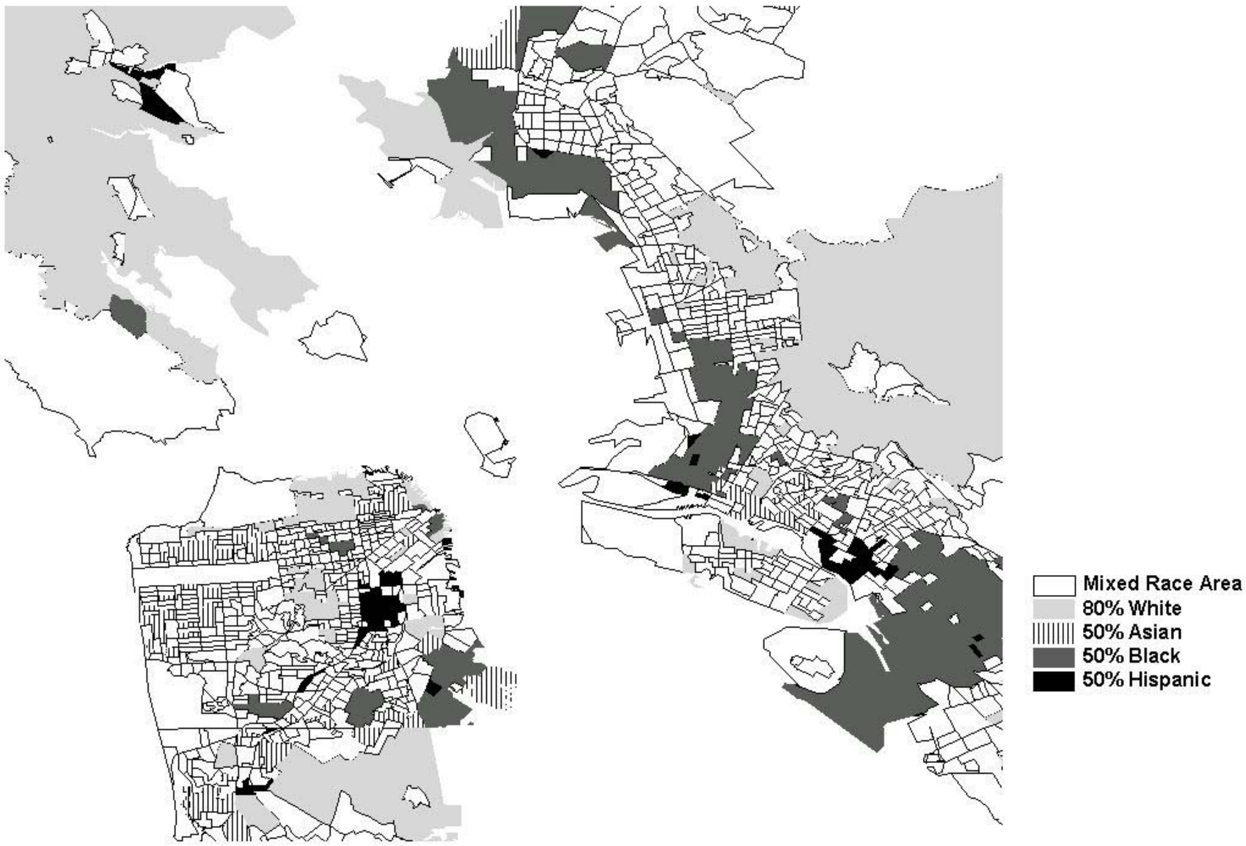
Schelling, Thomas C., (1971), “Dynamic Models of Segregation,” *Journal of Mathematical Sociology*, 1: 143-186 .

Schelling, Thomas C., (1978), *Micromotives and Macrobehavior*, Norton: New York.

Schnare, Ann, (1976), “Racial and Ethnic Price Differentials in an Urban Housing Market,” *Urban Studies*, 13: 107-120.

Yinger, John, (1978), “The Black-White Price Differential in Housing: Some Further Evidence,” *Land Economics*, 54:197-206.

Figure 1: Segregation Patterns in the Bay Area



Note: This figure provides a geographic depiction of segregation patterns for only the central portion of the full study area used in the analysis. San Francisco is the peninsula shown on the lower left of the figure; Oakland is located to the East of San Francisco directly across the Bay; Berkeley and Richmond are located North of Oakland in the upper right portion of the figure; and the upper left part of the figure shows a portion of Marin County.

Table 1: Racial Segregation in the San Francisco Bay Area

	Average Racial Composition of Census Block Group			
	Percent Asian	Percent Black	Percent Hispanic	Percent White
Household - Asian	22.5%	8.3%	11.7%	57.4%
Household - Black	11.6%	40.1%	11.4%	36.9%
Household - Hispanic Origin	12.9%	9.1%	21.8%	56.2%
Household - White	10.4%	4.8%	9.3%	75.5%
Overall Composition of Bay Area	12.3%	8.8%	11.2%	67.7%

	Asian	Black	Hispanic	White
Over-Exposure to Own Race	10.2%	31.3%	10.6%	7.8%

Note: Each of the first four rows shows the average racial composition of the block groups in which households of the race shown in the row heading reside. For comparison the fifth row shows the overall racial composition of the Bay Area. The 'Over-Exposure to Own Race' measure is defined for each race as the difference between the fraction of same-race neighbors (in same Census block group) and the overall fraction of households of the same race in the Bay Area.

Table 2: Mean Values of Selected Household Characteristics for Households of Each Race

Variable	Asian	Black	Hispanic	White	Overall
Household head is high school dropout	0.19	0.23	0.39	0.10	0.16
Household head graduated from high school	0.14	0.22	0.22	0.18	0.18
Household head has some college	0.18	0.28	0.19	0.23	0.23
Household head has bachelor's degree	0.33	0.20	0.16	0.32	0.29
Household head has advanced degree	0.16	0.06	0.05	0.17	0.14
Household income less than $12K	0.13	0.26	0.14	0.10	0.12
Household income $12-20K	0.09	0.14	0.12	0.08	0.09
Household income $20-35K	0.18	0.23	0.24	0.19	0.20
Household income $35-50K	0.18	0.16	0.21	0.18	0.18
Household income $50-75K	0.23	0.14	0.19	0.22	0.21
Household income $75-100K	0.12	0.05	0.07	0.11	0.10
Household income more than $100K	0.08	0.03	0.04	0.12	0.10
Household receives public assistance income	0.13	0.21	0.11	0.05	0.08
Household has capital gains or dividend income	0.48	0.17	0.25	0.56	0.48
Household head over 65	0.13	0.17	0.11	0.21	0.18
Household head divorced	0.07	0.20	0.14	0.16	0.15
Number of adults in the household	2.48	1.85	2.40	1.86	2.00
Number of pre-kindergarten children in household	0.31	0.27	0.40	0.17	0.22
Number of children grades K-8 in household	0.46	0.41	0.54	0.22	0.30
Number of children grades 9-12 in household	0.14	0.11	0.14	0.06	0.08
Spanish spoken in household	0.01	0.04	0.68	0.03	0.10
Asian language spoken in household	0.76	0.01	0.02	0.01	0.11
Household head born in US	0.24	0.97	0.54	0.90	0.78
Household head not a US citizen	0.35	0.02	0.31	0.04	0.11
Household head a naturalized citizen	0.41	0.01	0.15	0.06	0.11
Household head entered the US in 1980s	0.33	0.02	0.15	0.02	0.07
Household head entered the US in 1970s	0.26	0.01	0.14	0.02	0.06
Number of Observations	30271	18501	26675	167897	243344

Table 3: Stratification on the Basis of Educational Attainment in the San Francisco Bay Area

	Educational Attainment Distribution of Census Block Group				
	Percent HH High School Dropout	Percent HH with High School Degree	Percent HH with Some College	Percent HH with Bachelor's Degree	Percent HH with Advanced Degree
Household's Educational Attainment:					
HH - High School Dropout	25.7%	21.5%	21.9%	22.5%	8.4%
HH - High School Diploma	18.1%	21.3%	23.5%	26.3%	10.7%
HH - Some College	15.0%	19.1%	23.7%	29.1%	13.0%
HH - Bachelor's Degree	12.0%	16.6%	22.6%	32.1%	16.7%
HH - Advanced Degree	9.1%	13.7%	20.5%	33.9%	22.8%
Overall	15.5%	18.4%	22.6%	29.1%	14.4%

Notes: Each entry in the table shows the average fraction of households with the level of educational attainment shown in the column heading that reside in the same neighborhood as households with the level of education attainment shown in the row heading.

Table 4: Racial Exposure Rates by Race and Education

	Average Racial Composition of Census Block Group			
	% Asian	% Black	%Hispanic	%White
Household's Race, Education:				
Asian, No HS Diploma	30.7%	11.7%	13.4%	44.2%
Asian, HS Diploma	22.7%	10.1%	13.3%	54.0%
Asian, Some College	21.3%	8.9%	12.9%	56.9%
Asian, BA	20.8%	7.0%	11.4%	60.8%
Asian, More than BA	18.0%	4.9%	8.2%	68.9%
Black, No HS Diploma	10.3%	52.6%	11.4%	25.7%
Black, HS Diploma	11.0%	44.4%	12.2%	32.4%
Black, Some College	12.4%	36.0%	11.8%	39.8%
Black, BA	12.5%	30.3%	10.8%	46.5%
Black, More than BA	13.1%	25.7%	8.6%	52.7%
Hispanic, No HS Diploma	12.7%	11.1%	28.9%	47.4%
Hispanic, HS Diploma	13.1%	8.4%	21.6%	57.0%
Hispanic, Some College	13.2%	7.7%	17.2%	61.8%
Hispanic, BA	13.1%	7.6%	14.3%	65.1%
Hispanic, More than BA	12.5%	6.2%	10.5%	70.9%
White, No HS Diploma	11.0%	6.2%	13.5%	69.2%
White, HS Diploma	10.4%	5.0%	11.1%	73.5%
White, Some College	10.4%	4.8%	9.8%	75.1%
White, BA	10.4%	4.5%	8.1%	77.0%
White, More than BA	10.1%	4.1%	6.4%	79.4%
Overall	12.3%	8.8%	11.2%	67.7%

Notes: Each entry in the table shows the average fraction of households of the race shown in the column heading that reside in the same neighborhood as households of the race and level of educational attainment shown in the row heading.

Table 5: Calculating Racial Exposure Rate Measures Conditioning on Education

	Exposure to Households of Same Race (1)	Education Distribution *Own-Race* (2)	Education Distribution *Overall Sample* (3)	Contribution to Exposure Measure *Own-Race Educ Distribution* (4) = (1)*(2)	Contribution to Exposure Measure *Overall Sample Educ Distribution* (5) = (1)*(3)	Percent Reduction in 'Over-Exposure' to Households of Same Race (6)
Asian Households						
No HS Diploma	0.307	0.190	0.160	0.058	0.049	
HS Diploma	0.227	0.140	0.180	0.032	0.041	
Some College	0.213	0.180	0.230	0.038	0.049	
BA Degree	0.208	0.330	0.290	0.069	0.060	
More than BA	0.180	0.160	0.140	0.029	0.025	
				0.226	0.224	**1.36%**
Black Households						
No HS Diploma	0.526	0.230	0.160	0.121	0.084	
HS Diploma	0.444	0.220	0.180	0.098	0.080	
Some College	0.360	0.280	0.230	0.101	0.083	
BA Degree	0.303	0.200	0.290	0.061	0.088	
More than BA	0.257	0.060	0.140	0.015	0.036	
				0.395	0.371	**8.05%**
Hispanic Households						
No HS Diploma	0.289	0.390	0.160	0.113	0.046	
HS Diploma	0.216	0.220	0.180	0.048	0.039	
Some College	0.172	0.190	0.230	0.033	0.040	
BA Degree	0.143	0.160	0.290	0.023	0.041	
More than BA	0.105	0.050	0.140	0.005	0.015	
				0.221	0.181	**36.86%**
White Households						
No HS Diploma	0.692	0.100	0.160	0.069	0.111	
HS Diploma	0.735	0.180	0.180	0.132	0.132	
Some College	0.751	0.230	0.230	0.173	0.173	
BA Degree	0.770	0.320	0.290	0.246	0.223	
More than BA	0.794	0.170	0.140	0.135	0.111	
				0.756	0.750	**6.87%**

Notes: Column (1) reports own-race exposures by education calculated in Table 4.
Columns (2) and (3) are the education distributions for each race and the full sample respectively (see Table 2).
Columns (4) and (5) are calculated as listed in column heading, the last row for each panel is the sum of the rows above.
Column (6) is calculated as (Column (4) -Column (5))/(Column (4) - % of race in overall sample), thereby measuring how the over exposure to others of the same race is explained by the education distribution differences.

Table 6: Reduction in Percentage of Hispanic Segregation Related to Educational Attainment Holding Intensity of Exposure Constant

Panel A

	Average Exposure to Hispanic Hhlds in Educ Category							Repeating Analysis of Table 5				
	Hispanic <HS	Hispanic HS Deg	Hispanic Some Col	Hispanic BA	Hispanic > BA	Hispanic Total (1)		Hispanic Educ Distrib (2)	Overall Educ Distrib (3)	Calculating Average Exposure Measure (1)*(2)	(1)*(3)	
Hispanic Households												
No HS Diploma	0.165	0.057	0.038	0.024	0.005	0.289		0.390	0.160	0.113	0.046	
HS Diploma	0.104	0.051	0.034	0.022	0.005	0.216		0.220	0.180	0.048	0.039	
Some College	0.075	0.037	0.032	0.022	0.006	0.172		0.190	0.230	0.033	0.040	
BA Degree	0.057	0.030	0.026	0.024	0.006	0.143		0.160	0.290	0.023	0.041	**Reduction in 'Over-Exposure' to Hispanic Households *Fixed Exposure Rates***
More than BA	0.034	0.022	0.021	0.018	0.010	0.105		0.050	0.140	0.005	0.015	
Fraction of Total Bay Area Population:	0.044	0.025	0.021	0.018	0.006	0.112				**0.221**	**0.181**	**36.9%**

Panel B

	Intensity of Exposure Measures					
	Hispanic <HS	Hispanic HS Deg	Hispanic Some Col	Hispanic BA	Hispanic > BA	Hispanic Total
Hispanic Households						
No HS Diploma	3.776	2.317	1.784	1.341	0.893	2.580
HS Diploma	2.380	2.073	1.596	1.229	0.893	1.929
Some College	1.716	1.504	1.502	1.229	1.071	1.536
BA Degree	1.304	1.220	1.221	1.341	1.071	1.277
More than BA	0.778	0.894	0.986	1.006	1.786	0.938

Panel C

	Counterfactual: New Exposure Rates Adjusting Education Distribution									
	Hispanic <HS	Hispanic HS Deg	Hispanic Some Col	Hispanic BA	Hispanic > BA	Hispanic Total (1)		Overall Educ Distrib (2)	Calculating Exposure (1)*(2)	
Hispanic Households										
No HS Diploma	0.068	0.047	0.046	0.044	0.014	0.218		0.160	0.035	
HS Diploma	0.043	0.042	0.041	0.040	0.014	0.180		0.180	0.032	
Some College	0.031	0.030	0.039	0.040	0.017	0.157		0.230	0.036	
BA Degree	0.023	0.025	0.031	0.044	0.017	0.140		0.290	0.041	**Reduction in 'Over-Exposure' to Hispanic Households *Fixed Intensity of Exposure***
More than BA	0.014	0.018	0.025	0.033	0.028	0.118		0.140	0.017	
Fraction of Total Bay Area Population: Adjusting Education Distribution	0.018	0.020	0.026	0.033	0.016	0.112			**0.160**	**55.7%**

Notes: The rows in Panel A describe the average fraction of Hispanic households with the level of education listed in the column heading that reside in the Census block groups of the Hispanic households with the level of education listed in the row heading. For example, an average of 5.7% of the neighbors of a Hispanic household with a BA degree are Hispanic households with a less than a HS degree. The right-hand side of Panel A repeats the analysis of Table 5.

The rows in Panel B rewrite the exposure rates of Panel A as a percentage of the overall fraction of Hispanic households with the education shown in the column heading in the Bay Area. For example, Hispanic households with a HS degree live on average with twice as many Hispanic households with a HS degree as are represented in the Bay Area as a whole.

Using the intensity of exposure measures of Panel B, Panel C recalculates the exposure rate measures of Panel A under the counterfactual that the distribution of education for Hispanic households matched that of the full population of the Bay Area. The distribution of Hispanic households by education category that corresponds to this counterfactual is shown in the last row of Panel C. The right-hand side of Panel C calculates the overall own-race exposure of Hispanic households using the new exposure measures calculated on the left-hand side of Panel C.

Table 7: Explaining Exposure to Households of the Same Race

Dependent Variable:	% Asian	% Black	% Hispanic	% White
Sub-Sample:	Asian Hhlds	Black Hhlds	Hispanic Hhlds	White Hhlds
Observations	30,271	18,501	26,675	167,897
Adjusted R-squared	0.127	0.156	0.205	0.090
HH Education Level:				
No HS Diploma	0.054	0.118	0.099	-0.077
	(0.010)	(0.013)	(0.006)	(0.004)
HS Diploma	0.018	0.094	0.064	-0.038
	(0.005)	(0.012)	(0.005)	(0.003)
Some College	0.016	0.049	0.036	-0.021
	(0.004)	(0.011)	(0.005)	(0.002)
BA diploma	0.017	0.027	0.024	-0.010
	(0.004)	(0.010)	(0.004)	(0.002)
Household Income Information:				
< $12K	0.055	0.210	0.078	-0.105
	(0.012)	(0.017)	(0.008)	(0.004)
$12K-20K	0.028	0.189	0.066	-0.089
	(0.008)	(0.016)	(0.007)	(0.004)
$20-35K	0.015	0.147	0.063	-0.074
	(0.007)	(0.015)	(0.006)	(0.003)
$35-50K	0.017	0.116	0.046	-0.062
	(0.006)	(0.015)	(0.006)	(0.003)
$50-75K	0.027	0.085	0.033	-0.048
	(0.005)	(0.014)	(0.005)	(0.003)
$75-100K	0.014	0.038	0.014	-0.030
	(0.004)	(0.015)	(0.006)	(0.002)
Receives Public Assistance	0.002	0.053	0.019	-0.045
	(0.006)	(0.008)	(0.005)	(0.003)
Capital Gains or Dividend Income	0.005	-0.017	-0.018	0.008
	(0.002)	(0.007)	(0.003)	(0.001)
Language Spoken in Household:				
Spanish	-0.001	-0.036	0.051	-0.034
	(0.011)	(0.013)	(0.004)	(0.003)
Other European Language	0.011	-0.033	-0.001	-0.010
	(0.006)	(0.014)	(0.008)	(0.002)
Asian Language	0.048	-0.065	0.005	-0.075
	(0.004)	(0.020)	(0.008)	(0.005)
Other Language	0.005	0.005	-0.012	-0.033
	(0.022)	(0.019)	(0.017)	(0.006)
HH English Ability:				
Speaks English Well	0.004	0.006	0.010	-0.014
	(0.003)	(0.027)	(0.004)	(0.004)
Speaks Some English	0.025	0.038	0.034	-0.047
	(0.008)	(0.031)	(0.005)	(0.007)
Speaks No English	0.158	-0.138	0.055	-0.082
	(0.031)	(0.085)	(0.011)	(0.020)
HH Citizenship Status:				
Not Citizen	0.024	-0.059	0.016	0.012
	(0.006)	(0.037)	(0.008)	(0.006)
Naturalized Citizen	0.033	-0.031	0.003	0.003
	(0.006)	(0.035)	(0.007)	(0.005)
Entered Country in 1980’s	-0.024	-0.067	-0.026	-0.024
	(0.008)	(0.036)	(0.009)	(0.007)
Entered Country in 1970’s	-0.002	-0.060	-0.012	-0.021
	(0.008)	(0.036)	(0.008)	(0.006)
Entered Country pre-1970	-0.005	-0.089	-0.021	-0.005
	(0.007)	(0.031)	(0.007)	(0.005)
Constant	0.093	-0.066	0.023	0.798
	(0.020)	(0.036)	(0.020)	(0.012)

Notes: Each column shows the results of regressing the fraction of households of the race shown in the column heading on the set of household characteristics shown in the rows using only the sub-sample of households of the same race. The regressions also control for marital status and age of householder, number of adults and children in household, military service history of household and ten broad occupation categories for householder. Omitted categorical variables for each set of regressors are: more than a BA for education, income over $100K, speaks only English, speaks English very well, and native born.

Table 8: Reduction of Racial Segregation Explained By Non-Racial Household Characteristics

	Operation	Asian	Black	Hispanic	White
Baseline:					
(1) Representation of Race in SF Bay Area	(1)	12.3%	8.8%	11.2%	67.7%
(2) Exposure to Households of Same Race	(2)	22.5%	40.0%	21.8%	75.5%
(3) "Over-Exposure" to Households of Same Race	(3) = (2) - (1)	10.2%	31.2%	10.6%	7.8%
Controlling for Full Set of Household Characteristics:					
(4) "Conditional Exposure" to Households of Same Race	(4)	17.1%	30.5%	12.0%	73.0%
(5) Percentage Point Decline in Exposure Rate	(5) = (2) - (4)	5.5%	9.4%	9.8%	2.5%
(6) Amount Explained by Household Characteristics	(6) = (5)/(3)	**53.2%**	**30.3%**	**92.5%**	**32.4%**

Household Characteristics	Percentage Reduction in Exposure to Households of Same Race			
	Asian	Black	Hispanic	White
Educational Attainment	**0.8%**	**4.3%**	**19.3%**	**4.9%**
Total Effect of Income	**0.7%**	**14.2%**	**10.2%**	**6.6%**
Income Level	0.7%	10.2%	5.6%	3.9%
Household on Public Assistance Income	0.1%	2.3%	0.5%	1.8%
Has Non-Salary Wealth	0.0%	1.7%	4.0%	0.8%
Household Language Effects	**38.7%**	**3.0%**	**32.3%**	**15.2%**
Non-English Language Spoken	30.3%	3.1%	27.4%	11.7%
English Ability	8.3%	-0.1%	4.8%	3.5%
Total Citizenship Effect	**8.5%**	**6.9%**	**-1.7%**	**1.7%**
Citizen Status	15.2%	2.5%	3.2%	-1.2%
Years in US	-6.8%	4.4%	-4.9%	3.0%
Household Demographics	**1.3%**	**0.3%**	**13.9%**	**1.7%**
Military Service	0.8%	-0.3%	0.8%	-0.3%
Occupation	1.1%	1.6%	4.0%	0.8%
Total	**53.2%**	**30.3%**	**92.5%**	**32.4%**

Notes : Rows (1) - (3) correspond to the exposure rate measures described in Table 1. Row (4) presents the fraction of households of the same race in the neighborhood predicted using the regression coefficients in Table 7 for each race and the overall population means for the full set of household characteristics included on the right-hand side of these regressions. Rows (5) and (6) present the corresponding predicted decline in own-race 'over-exposure'.

The lower panel decomposes the calculated decline in own-race 'over-expsoure' associated with the particular set of household characteristics listed in the row heading. These values are based on predicted exposure rates obtained using the regression coefficients for each race in Table 7, replacing each race's own mean for the set of household characteristics listedin the row heading with the overall mean for the Bay Area population.

Table 9: Exploring Reductions in Residential Segregation Under Different Exposure Scenarios

	Number of Distinct Race-Hhld Characteristic Categories	Percentage Reduction in Exposure Households of the Same Race			
		Asian	Black	Hispanic	White
Educational Attainment Only					
Fixed Exposure Rates	20	1.4%	8.1%	36.9%	6.9%
Fixed Intensity of Exposure	20	2.1%	16.2%	55.7%	10.7%
Education (2) x Income (3) x Language (2)					
Fixed Exposure Rates	48	44.6%	20.1%	61.9%	9.8%
Fixed Intensity of Exposure	48	62.1%	33.2%	78.5%	30.3%
Education (2) x Income (3) x Immigrant Status (2)					
Fixed Exposure Rates	48	19.6%	22.5%	39.6%	21.6%
Fixed Intensity of Exposure	48	21.2%	36.6%	59.1%	22.7%
Education (2) x Income (3) x Public Assistance Income (2)					
Fixed Exposure Rates	48	0.3%	16.9%	30.6%	9.0%
Fixed Intensity of Exposure	48	0.7%	25.8%	50.1%	12.7%
Education (2) x Public Assistance Income (2) x Language (2) x Immigration Status (2)					
Fixed Exposure Rates	64	46.3%	19.6%	57.4%	14.3%
Fixed Intensity of Exposure	64	67.8%	31.2%	76.2%	24.8%

Notes: This table presents the results from several conterfactuals. The rows labeled 'Fixed Exposure Rates' report the results from counterfactuals that treat exposure rates as primitives, while the rows labeled 'Fixed Intensity of Exposure ' report the results from counterfactuals that treat intensity of exposure measures as primitives. Each panel uses interactions of race with the distinct categories of household characteristics shown in each row heading. The education categores distinguish households that have received at least a bachelor's degree; the income categories distinguish: less than $35k, $35-75k, and $75k+; the language categories distinguish those that speak a foreign language; the immigration status categories distinguish native-born US citizens; and the public assistance categories distinguish those receiving any form of public assistance income.

Appendix Table 1: Racial Segregation in the Bay Area at the Census Block Group Level

Average Racial Composition of Census Block Group by Race of Household

	Racial Composition of Census Block Group (1)			
	Percent Asian	Percent Black	Percent Hispanic	Percent White
Asian Household	22.5%	8.3%	11.7%	57.4%
Black Household	11.6%	40.1%	11.4%	36.9%
Hispanic Household	12.9%	9.1%	21.8%	56.2%
White Household	10.4%	4.8%	9.3%	75.5%
Overall	12.3%	8.8%	11.2%	67.7%

Underlying Coefficient Estimates and Standard Errors (2)

	Percent Asian	Percent Black	Percent Hispanic	Percent White
Asian Household	0.121	0.035	0.024	-0.181
	(0.008)	(0.003)	(0.004)	(0.009)
Black Household	0.012	0.353	0.021	-0.386
	(0.004)	(0.015)	(0.004)	(0.013)
Hispanic Household	0.025	0.043	0.125	-0.193
	(0.004)	(0.003)	(0.009)	(0.010)
Constant	0.104	0.048	0.093	0.755
	(0.003)	(0.002)	(0.002)	(0.005)
Observations	243419	243419	243419	243419
Adjusted R-squared	0.108	0.337	0.112	0.273

Notes:
(1) The exposure rates in the top panel are calculated by adding the own race coefficient in the bottom panel to the constant in each column.
(2) For exposure rates at different levels of aggregation, see Appendix Tables 2a and 2b.

Appendix Table 2:
Racial Exposure Rates at Different Levels Of Aggregation

	Census Block Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	26.1%	7.7%	11.2%	55.0%
	(0.009)	(0.003)	(0.003)	(0.008)
Black Household	11.2%	42.8%	11.2%	34.9%
	(0.004)	(0.016)	(0.004)	(0.013)
Hispanic Household	12.5%	8.8%	24.9%	53.8%
	(0.003)	(0.003)	(0.008)	(0.008)
White Household	10.0%	4.4%	8.8%	76.8%
	(0.002)	(0.002)	(0.002)	(0.004)
Observations	242218	242218	242218	242218
Adjusted R-squared	0.081	0.288	0.074	0.21

	Census Block Group Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	22.5%	8.3%	11.7%	57.4%
	(0.008)	(0.003)	(0.004)	(0.009)
Black Household	11.6%	40.1%	11.4%	36.9%
	(0.004)	(0.015)	(0.004)	(0.013)
Hispanic Household	12.9%	9.1%	21.8%	56.2%
	(0.004)	(0.003)	(0.009)	(0.010)
White Household	10.4%	4.8%	9.3%	75.5%
	(0.003)	(0.002)	(0.002)	(0.005)
Observations	243419	243419	243419	243419
Adjusted R-squared	0.108	0.337	0.112	0.273

	Census Tract Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	21.4%	8.5%	11.9%	58.4%
	(0.008)	(0.003)	(0.004)	(0.008)
Black Household	11.8%	38.3%	11.7%	38.2%
	(0.005)	(0.020)	(0.004)	(0.016)
Hispanic Household	13.1%	9.3%	20.8%	57.0%
	(0.003)	(0.004)	(0.008)	(0.009)
White Household	10.6%	5.0%	9.5%	75.0%
	(0.003)	(0.002)	(0.002)	(0.005)
Observations	243422	243422	243422	243422
Adjusted R-squared	0.105	0.329	0.111	0.27

	PUMA Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	16.2%	9.1%	12.1%	62.6%
	(0.009)	(0.013)	(0.010)	(0.021)
Black Household	12.9%	25.6%	12.1%	49.4%
	(0.012)	(0.062)	(0.009)	(0.062)
Hispanic Household	13.4%	9.4%	15.7%	61.5%
	(0.010)	(0.014)	(0.016)	(0.026)
White Household	11.5%	6.4%	10.2%	71.9%
	(0.010)	(0.011)	(0.009)	(0.020)
Observations	243425	243425	243425	243425
Adjusted R-squared	0.053	0.194	0.062	0.168

	County Racial Composition			
	Percent Asian	Pecent Black	Percent Hispanic	Percent White
Asian Household	13.9%	8.9%	11.6%	65.6%
	(0.012)	(0.005)	(0.006)	(0.018)
Black Household	12.5%	12.4%	10.7%	64.4%
	(0.008)	(0.013)	(0.008)	(0.016)
Hispanic Household	12.7%	8.4%	11.9%	67.0%
	(0.006)	(0.006)	(0.004)	(0.014)
White Household	11.9%	8.4%	11.1%	68.6%
	(0.018)	(0.027)	(0.013)	(0.033)
Observations	243425	243425	243425	243425
Adjusted R-squared	0.022	0.04	0.013	0.034

Notes:
Standard errors in parentheses.

Data Appendix

This data appendix gives descriptions of and summary statistics on all the variables used in the analysis.
The following summary statistics are based on a sample of 243,350 households drawn from the 6 Bay Area counties.
Person weights drawn from the Census are used when calculating the household and neighborhood level numbers.

Variable Description	Mean	Std. Dev.
household head is high school dropout	0.16	0.36
household head graduated from high school	0.18	0.39
household head has some college	0.23	0.42
household head has bachelor’s degree	0.29	0.45
household income less than $12K	0.12	0.32
household income $12-20K	0.09	0.29
household income $20-35K	0.20	0.40
household income $35-50K	0.18	0.39
household income $50-75K	0.21	0.41
household income $75-100K	0.10	0.30
household receives public assistance income	0.08	0.27
household has dividend income	0.48	0.50
sex of household head	1.34	0.47
age of household head	46.98	16.63
household head over 65	0.18	0.39
household head widowed	0.10	0.30
household head divorced	0.15	0.35
household head separated	0.03	0.17
household head never married	0.21	0.41
number of adults in the household	2.00	0.98
number of pre-kindergarten children in household	0.22	0.56
number of children grades K-8 in household	0.30	0.70
number of children grades 9-12 in household	0.08	0.31
Spanish spoken in household	0.10	0.30
Asian language spoken in household	0.11	0.31
other European language spoken in household	0.07	0.26
other language spoken in household	0.01	0.09
household head speaks English well	0.06	0.24
household head speaks some English	0.04	0.19
household head speaks no English	0.01	0.09
household head not a US citizen	0.11	0.31
household head a naturalized citizen	0.11	0.31
household head entered the US in 1980s	0.07	0.26
household head entered the US in 1970s	0.06	0.24
household head entered US pre-1970	0.09	0.29
household head active in military	0.01	0.07
household head previously in military	0.22	0.41
household head in reserves	0.02	0.15